Story Telling Exercise: Doctor's Helper

Introduction

Doctor's Helper is an initiative started by an innovative doctor at a local hospital. After becoming overwhelmed with the volume of patients that she was seeing, the doctor realized that she needed to augment her capability with the latest advancement in data analysis and data science in order to help her find those patients that are at the highest risk for disease.

She seeks to enlist the help of a tech savvy analyst that may aid her quest to uncover the biological markers and environmental circumstances that may be associated with higher/lower incidence of disease.

The team starts with a small data set, compiled by the doctor, that lists the biomarkers and environmental factors that she feels are potentially important along with the binary variable "Outcome" indicating if the disease was present in the patient.

Directions

You are the Data Scientist working with the good doctor. Your job is to:

- Analyze the dataset provided by the doctor and deliver your insights.
- Fit an appropriate model to predict Outcome and evaluate the performance of the model.
- Provide recommendations to the doctor that will help discover patients at risk.
- Prepare a presentation to hospital staff that includes the result of your analysis. Your presentation may be as simple as creating a commented and clean Jupyter Notebook with code and charts that can walk the hospital staff through your findings).
- Remember that the hospital staff may not understand some of the analytical nuance. You have to present your analysis in a way that they can understand.
- Be prepared to give a 10-minute presentation during your interview and answer questions about your approach and analysis.

Dataset

The dataset provided contains 9 variables that are biological in nature and 2 that are environmental. The biological variables are:

- Number of Pregnancies
- Blood Chemistry I
- Blood Chemistry II
- Blood Chemistry III
- Blood Pressure
- Skin Condition
- BMI
- Genetic Predisposition Factor
- Age

The Environmental variables are:

- Air Quality Index
- Geographic Location (State)